Population Issues in Welfare Economics, Ethics, and Policy Evaluation

Kevin Kuruc, Department of Economics, University of Oklahoma, Mark Budolfson, Department of Environmental and Occupational Health and Justice, Center for Population-Level Bioethics, and Department of Philosophy, Rutgers University, and Dean Spears, Department of Economics and Population Research Center, University of Texas at Austin; IZA; Economics and Planning Unit, Indian Statistical Institute

https://doi.org/10.1093/acrefore/9780190625979.013.727
Published online: 24 February 2022

Summary

Nearly all large policy decisions influence not only the quality of life for existing individuals but also the number—and even identities—of yet-to-exist individuals. Accounting for these effects in a policy evaluation framework requires taking difficult stances on concepts such as the value of existence. These issues are at the heart of a literature that sits between welfare economics and philosophical population ethics. Despite the inherent challenges of these questions, this literature has produced theoretical insights and subsequent progress on variable-population welfare criteria. A surprisingly bounded set of coherent alternatives exists for practitioners once a set of uncontroversial axioms is adopted from the better-studied welfare criteria for cases where populations are assumed to be fixed. Although consensus has not yet been reached among these remaining alternatives, their recommendations often agree. The space has been sufficiently restricted and well explored that applications of the theoretical insights are possible and underway in earnest.

For reasons both theoretical and empirical, the applied literature studying optimal policy and its robustness to welfare criteria has documented a surprising degree of convergence between recommendations under quite different ethical stances on existence value. This convergence has appeared even in cases where population size itself is the choice variable. Although it may not always be the case that policy recommendations are invariant to population welfare criteria, tools have been developed that allow researchers to easily and transparently move between such criteria to study the robustness in their context of interest. The broader point is that the remaining theoretical uncertainties need not prevent population ethics from playing a role in policy evaluation—the tools are available for determining whether and which policies are broadly supported among a range of ethical views.

Keywords: population, welfare economics, public choice, long-run development, climate change

Subjects: Economic Development, Health, Education, and Welfare Economics, Labor and Demographic Economics

Population Ethics and Welfare Economics

Many socioeconomic policies impact both the number and identity of future people. Writing of the tragic Chinese famine of the late 1950s, Angus Deaton mourned the actual deaths as well as the lives never lived because of this event: “One of the worst in human history was China’s ‘Great Leap Forward’ in 1958–1961, when deeply misguided industrialization and food procurement policies led to the deaths of around thirty-five million people from starvation and prevented the births of perhaps forty million more” (Deaton, 2013, p. 38). Economists have tools to evaluate the social welfare loss incurred from the deaths of the 35 million, although debates about the details continue, to be sure. An apparently open question is how to value the absence of the 40 million? Do they count among the costs of this catastrophe?

In a contemporary example, climate change may increase early-life mortality or decrease the carrying capacity of the earth, hence indirectly affecting population dynamics into the far future. How, if at all, should lives not lived enter policy evaluation? It may be justifiable to ignore this question when population impacts are second-order, hard-to-forecast effects of an already-bad event. However, social choices where population size is predictably affected—or is itself the choice variable—are not rare. Should fertility be reduced to alleviate future environmental pressures? Is it welfare enhancing to shrink the factory farming industry when the alternative for the animals is nonexistence? How much should be spent to reduce the probability of human extinction? How should the decline in the size of the world population that economists project for the 22nd century be evaluated?

These questions pose unique challenges relative to conventional economic policy options where the number of individuals is assumed to be fixed, that is, fixed population cases. When populations are held fixed, maximizing total (summed) well-being is equivalent to maximizing average well-being conditional on existence, which are both equivalent to maximizing the well-being of only the individuals who exist in all outcomes. When future populations are non-constant—variable population cases—these three objectives can lead to divergent recommendations. Consider a fictitious decision that impacts no currently existing person, but brings one new individual into existence with positive, but below average, well-being: this increases total well-being, decreases average well-being conditional on existence, and is neutral on the metric that considers only “always existing” individuals. In this intentionally simple case, three distinct evaluations are obtained.

In an effort to provide practitioners and applied economists a framework for evaluating decisions that impact the size and composition of future populations, this review article explores the intersection of population ethics—which is the philosophical literature on social choice when the number and identity of individuals varies—and welfare economics. It begins by discussing theoretical advances that extend concepts of Pareto-like efficiency and aggregative methods for social welfare to variable population cases. With the theoretical landscape covered, the practical implications for economic decisions and a generalized social welfare function are examined.

Despite the inherent challenges associated with valuing existence, impressive theoretical progress has been made on these problems. The current state of knowledge can be summarized in the following way. Traditional concepts of efficiency become quickly useless when socioeconomic decisions influence who exists; this is thought to be the case for a surprisingly large class of decisions (Broome, 2018). In part because of the lack of guidance efficiency concepts then provide, attention has been directed at constructing aggregative measures of social welfare. Starting with a small set of uncontroversial axioms leads to surprisingly binding restrictions on the class of feasible social welfare functions (SWFs). Furthermore, the resulting SWFs in many cases challenge widely held intuitions; at present, researchers are left with the decision either to reject seemingly uncontroversial axioms or, as is recommended here, to move beyond familiar but often unexamined intuitions about how to evaluate welfare in populations.

To preview one such result: Broome (2004) showed that accepting even just three axioms that are unrelated to existence value (completeness, transitivity, and same-population Pareto) leads to the result that variable population social orderings must respond positively (negatively) to an additional (missing) person, given that person has (would have had) lifetime utility exceeding some level. Many people are initially “in favor of making people happy, but neutral about making happy people” (Narveson, 1973, p. 80). Discovering that no plausible social ordering supports this intuition forces one to abandon either this intuition or the concept of a complete, transitive, same-people-Pareto SWF. Layering additional axioms into this framework has produced similar dilemmas that clarify which combinations of intuitions and primitives are (in)coherent, though disagreement remains over the most reasonable of the coherent combinations.

In part because of these unresolved theoretical considerations, applied research incorporating these concepts developed more slowly, yielding a new set of interesting and sometimes surprising findings. This work has uncovered a high degree of convergence among policy recommendations under social welfare functions that treat existence in distinct ways, even in cases where population sizes are predictably affected. Although these normative frameworks are opposed in unrealistic cases constructed to elicit disagreement—such as bringing a new individual into existence without changing the welfare of any already-existing person—the empirical facts seem to be such that these theories mostly agree on policy questions. It is becoming acknowledged that the remaining theoretical ethical disagreements need not prevent population ethics from contributing to applied research. In cases where a broad range of values delivers the same policy recommendations, the underlying ethical disagreements are not pivotal.

This pattern of convergence is underexplored and will surely fail to hold for some socioeconomic policies. The article concludes with a brief overview of standard practice for bringing population ethics into an applied framework so that practitioners can independently study the robustness of policy evaluations in their setting. One straightforward and popular approach is to evaluate a policy under the two most straightforward welfare functions that treat population sizes in opposite ways. For more comprehensive robustness exercises, the field has produced a general SWF with parameters governing the ethical disagreements

thought to be most relevant for social policy. The ethical uncertainty can then be dealt with in a similar manner to standard parametric uncertainty: researchers can transparently study which (if any) outcomes are robustly recommended under a broad combination of ethical parameter choices. Coupled with the first-order impact on welfare analysis that population ethics can have, these newly accessible tools can be expected to spur a generation of research less concerned with the remaining ethical dilemmas and more with rigorously evaluating the important socioeconomic decisions the global community faces.

Theoretical Welfare Considerations in Variable Population Settings

Cases where the number, identity, and welfare of individuals vary are considered. It is assumed that each individual's well-being can be summarized by a scalar representing their lifetime utility. Generally, different states of affairs are denoted as state (or "outcome") X of N people with lifetime utilities $\mathbf{W} = \{w_1, w_2, \ldots, w_N\}$ and state Y of M people with lifetime utilities $\mathbf{V} = \{v_1, v_2, \ldots, v_M\}$. When it is relevant, the identity of the individuals is specified.

Efficiency Concepts and (Non-)Identity Problems

The concepts of efficiency and pairwise dominance are first examined. In standard same-number, same-identity cases, Pareto efficiency identifies all states where no individual could be made better off without harming any other. Likewise, Pareto dominance occurs when some state is weakly better for all individuals. What is attractive about the dominance relation is that it allows for normative statements without interpersonal comparisons—Pareto improvements are thought to be unambiguously (social) welfare improving because every individual prefers the new outcome. However, in variable-number, variable-identity cases, not all individuals exist in all outcomes. This makes it difficult to conceptualize what it means for an individual to prefer an outcome and greatly limits the usefulness of such concepts outside of same-identity cases.

Golosov et al. (2007) (henceforth GJT) was a landmark study of efficiency and dominance in settings of variable populations and clearly lays out the issues at hand. Its findings can be best presented in a stylized example that is returned to throughout the article. Assume that state X and state Y are identical with the exception that state Y has one additional individual with lifetime utility v_M; all other individuals' lifetime utilities are equal across states. Determining whether state Y dominates, in the sense that all individuals are as well off in state Y as in X, hinges on whether the additional person existing in Y is better off than when not existing in X. Economists who are accustomed to evaluating welfare based on preferences will wonder whether the additional person in state Y can be said to prefer existence to nonexistence. GJT considered two options for representing nonexistence: (a) as undefined in the lifetime utility space and so wholly incomparable to existence or (b) as a numerical utility level $(\underline{u})$ that can be compared with existence. To conceptualize a numerical utility level of nonexistence, $\underline{u}$, consider a life certainly worth living (perhaps your own) and a life you consider not worth

living (perhaps a short life of constant abuse); if the quality of your own life is continuously decreased toward this short, abused life, at some point there will be a threshold where existence is net-neutral, $\underline{u}$.

Rather than defend either of these as the proper conception of nonexistence, GJT presented two alternative concepts of efficiency or dominance corresponding to these competing definitions. What they call *A*-efficiency treats the utility of the nonexistent as undefined; *P*-efficiency treats the utility of the nonexistent as the numerical $\underline{u}$. *A*-efficiency, because it is premised on the fact that newly existing people are not made better off via existence, is concerned only with those individuals who exist in both states under consideration. In the stylized example where one new individual is added to the population without affecting any other individuals, *X* and *Y* weakly *A*-dominate each other. This is because state *Y*, with the additional person, makes no individuals better off if this newly existing person cannot be said to have benefited, nor is anyone harmed. Alternatively, *P*-efficiency claims that the newly existing person *is* made better off when brought into existence under the assumption that $v_M > \underline{u}$. Therefore, in this case state *Y* *P*-dominates state *X*. No individuals are harmed, and the individual coming into existence is made better off.

GJT discussed the implications and weaknesses of these competing concepts, which were furthered in Broome (2018). First, under the concept of *P*-efficiency it is impossible to have a population level that is ever too large as long as all newly existing individuals experience lifetime utility that exceeds nonexistence, even if the larger population outcome has lower average welfare. In any such pairwise decision situation, choosing the outcome with fewer individuals harms those who's existences are prevented, and hence cannot *P*-dominate the larger population outcome because dominance requires that *every* individual is made better off. This cuts to a core issue within population ethics: If the experienced utility of potential people in the states in which they exist is valued, large populations will have much in their favor.

Conversely, large populations will not be *A*-efficient if the additional individuals reduce the welfare of *any* individual that exists in both states. However, because *A*-relations require comparing only those alive in both states of interest, they can be shown not to satisfy transitivity. As an example, Table 1 presents some states X, Y, and Z such that X *A*-dominates Y, Y *A*-dominates Z, and Z *A*-dominates X (where numbers represent lifetime utilities; "-" represents nonexistence in *A*-relations).

Table 1. A-Efficiency Is Non-Transitive

	Person 1	Person 2	Person 3
X	10	10	–
Y	9	–	9
Z	–	11	0

A deeper issue with notions of efficiency in variable-population cases recognized in Broome (2018) surrounds empirical considerations regarding personal identity, on which dominance relations rely. Recall, the attractive feature of these relations is that it can be said whether all stake-holding individuals are weakly better off. This notion runs headfirst into what is known in the philosophical literature as the "non-identity problem" (Parfit, 1984). The non-identity problem highlights the difficulty of this style of reasoning under empirically realistic views of identity: namely, that if any of these respective conceptions occurred even just moments later, a different individual would have been brought into existence. Varying the sperm–egg combination of the same parental units results in a different person. If you doubt this, consider whether you would be your younger sibling in the counterfactual where your parents did not conceive you. It is hard to believe that would count in an ethically relevant sense as "the same 'you.'"

This view of identity is widely accepted in the philosophical literature and causes difficulties for efficiency concepts. It is surely the case that even conventional economic policy influences the identity of not-yet-conceived individuals; a large fiscal stimulus bill, for example, if it influences individual behavior at all, will change the sperm–egg combinations of future conceptions and hence, eventually, the entire set of future citizens. *A*-efficiency then, by considering only individuals who exist in both outcomes, essentially considers only the well-being of the already or soon-to-be conceived. *P*-efficiency, under this view of personal identity, would be exhaustive; any policy influences the identity of at least one future person, who would be harmed by this decision not being made. One is left either, in the case of *A*-efficiency, only considering already-conceived individuals, or, in the case of *P*-efficiency, scoring nearly every possible outcome as efficient. Neither is satisfactorily action guiding, and hence little can be said without interpersonal comparisons that exclude identity considerations (Broome, 2018).

Variable Population Social Welfare Functions

The study of the properties and classes of variable-population SWFs has made substantive progress since the problem first received attention. This work has produced famous incompatibility results (Arrhenius, 2000; Ng, 1989) that provide clarity on the ethical axioms that can be made to consistently fit together. Where they cannot, a rejection of one or more of these axioms guides the SWFs available to the practitioner.

Attention here is focused on preference relations that satisfy axioms that are uncontroversial in conventional (fixed-population) welfare economics.

1. (Social order) The social ordering is complete, transitive, and reflexive.
2. (Anonymity) For some vector of lifetime utilities, $\mathbf{v}$, and all vectors $\mathbf{w}$ that have identical entries to $\mathbf{v}$ but assign them to different individuals, the social ordering is indifferent between $\mathbf{v}$ and $\mathbf{w}$.
3. (Continuity) The social ordering is continuous.

4. (Same-number Pareto) For all $\mathbf{v}$ and $\mathbf{w}$ of equal length such that $\mathbf{v} > \mathbf{w}$, $\mathbf{v}$ is preferred to $\mathbf{w}$.

The existence of a complete, transitive, and continuous social order (axioms 1 and 3) ensures the existence of a SWF. The anonymity axiom says that the SWF is invariant to the rearrangement of utility values across people; in other words, the SWF cannot care *who* has the better or worse lives in cases where the entire distribution of utilities is fixed. Axiom 4 is merely the Pareto axiom for same number cases; if the SWF is ranking two outcomes with the same number of people and each individual prefers X to Y, the SWF must also prefer X to Y.

Even just these axioms produce an important result: if accepted, what has been called the "intuition of neutrality" cannot be correct (Broome, 2004). This intuition states that there exists a range of lifetime utilities such that all else equal, the existence of a new person with such a utility level is socially neutral. For example, it seems natural to assume that the world with or without the existence of a new individual is equally good, assuming this new individual does not have a sufficiently bad life. In policy debates this seems to be an underlying assumption behind calls for reduced populations: the proposed nonexistence of some people is not itself considered to have a direct effect on welfare. However, same-number Pareto, completeness, and transitivity preclude this assumption.

Proof: If there are three outcomes, X, Y, and Z, where Y and Z have an additional person whose life is better in Z than Y, but within the proposed "range of neutrality" in both, then $\mathbf{X} \sim \mathbf{Y}$ (by assumed neutrality); $\mathbf{X} \sim \mathbf{Z}$ (by assumed neutrality); but $\mathbf{Z} \succ \mathbf{Y}$ (by same-number Pareto). Transitivity is then violated. Therefore, X cannot be welfare equivalent to both Y and Z.

If, instead, completeness is rejected to retain this intuition (i.e., that X cannot be compared to Y or Z), essentially no policy can ever be stated to have any (social) welfare benefits or costs. At least indirectly, any policy will change the number of people who ever exist. Conceptually rejecting comparisons of outcomes with different population sizes leaves one unable to rank any policy changes with this property, even for policies where everyone who ever lives benefits. For a stark example, a rejection of completeness implies non-comparison between the current trajectory and one where humanity goes extinct tomorrow. This is a welfare comparison across outcomes with different numbers of ever-existing people; it seems one must be able to make statements that one outcome is better than the other.

Beyond "non-neutrality," axioms (1)–(4) guarantee that the social ordering can be represented as a function of two variables: the number of individuals, N, and the equally distributed equivalent (EDE) measure of individual welfare, Θ (Blackorby & Donaldson, 1984). The EDE is an inequality-adjusted average, such that if all individuals had this level of welfare, it would be as good as the actual distribution of welfare. In this article, ethical questions about inequality are ignored for the moment to focus on population issues, so Θ can be considered the arithmetic average of well-being.

To illustrate the form of these welfare functions, the two most well known and straightforward correspond to average utilitarianism (AU) (sometimes called a Millian SWF given John Stuart Mills's early implicit endorsement), and total utilitarianism (TU) (sometimes called a Benthamite SWF given Jeremey Bentham's implicit endorsement). These can be represented, respectively, as follows, where i indexes individuals and u_i is the agent's lifetime well-being:

$$V^{AU}(N,\Theta)= \frac{1}{N}\sum_{i=1}^{N} u_i = \Theta$$

$$V^{TU}(N,\Theta)= \sum_{i=1}^{N} u_i = N \times \Theta$$

AU, as suggested by the name, is increasing in the average well-being of a population, regardless of the population size; TU increases in both population size and average well-being as it is the sum of all experienced well-being. Because all variable population SWFs that satisfy axioms (1)–(4) are representable as functions of these two variables, nearly all SWFs proposed in the literature exhibit some trade-off between the quantity and average quality of lives, a trade-off that may or may not be constant. However, even this leaves open a considerable number of possibilities. A further narrowing of the set of feasible welfare functions is attainable through stances on a mere addition axiom and/or various independence axioms.

Mere Addition, Independence of the Utilities of the Dead, and Separability

***Mere addition*:** Holding all other individual's lifetime utility constant, the existence of a new individual with lifetime utility exceeding neutrality $(\underline{u})$ cannot reduce social welfare.

Mere addition is perhaps the weakest existence-related claim that can be added to the four same-number axioms above. It does not require that adding an individual with a net-pleasurable life is socially good—a less obvious claim—only that it cannot be socially bad. Many find this hard to reject. However, despite not being controversial on its own, when paired with axioms (1)–(4), the acceptance of mere addition implies the stronger claim that adding an individual with a net-pleasurable life is in fact strictly welfare increasing, other things equal (Roberts, 2020).

Proof: Suppose one has two states X and Y where X is only different from Y through the existence of an additional individual with utility level $u_{new} > \underline{u}$, that is, a net-pleasurable life. Z is the same as Y, only this new individual instead has lifetime utility $u'_{new} = u_{new} + \in$. By same-number Pareto, Z is strictly preferred to Y; the same individuals exist in Z and Y, with one individual being strictly better off. By mere addition, Y is not worse than (or, equivalently, weakly preferred to) X; a new individual exists with a net-pleasurable life. By transitivity then,

Z is strictly preferred to X. Z only differs from X in that it has one additional person with a net-pleasurable life. Therefore, mere addition plus axioms (1)–(4) imply that adding an individual with a net-pleasurable life is strictly welfare increasing, other things equal.

Further, when a non-anti-egalitarian axiom is included (i.e., that society doesn't strictly prefer inequality), total utilitarian population principles are the only principles that satisfy mere addition and the original four axioms.[1] However, total utilitarianism strikes many as unappealing because of a particular implication that its constant positive weight on all net-pleasurable lives has: For any given population, there is always a large population of barely net-pleasurable lives that is preferred (Parfit, 1984). This has come to be known as the repugnant conclusion, and seeking to avoid it dominated much of the population ethics literature for its first few decades.

Incompatibility theorems (Arrhenius, 2000; Ng, 1989) have shown that avoidance of the repugnant conclusion is incompatible with axioms (1)–(4), a non-anti-egalitarian axiom, and mere addition. This is exactly because totalist population principles are the only principles satisfying all axioms jointly, and these generate the repugnant conclusion. It is conceptually straightforward why mere addition leads quickly to the repugnant conclusion: If the addition of an individual with a net-pleasurable life is always welfare increasing (mere addition + completeness + transitivity), then the addition of many net-pleasurable lives can result in any arbitrarily large level of social welfare. So, efforts to avoid the repugnant conclusion but retain the concept of a SWF primarily run through a rejection of mere addition.

Blackorby et al. (1995) argued for acceptance of an axiom in place of mere addition that they referred to as "independence of the utilities of the dead." This axiom states only that rankings between outcome X and outcome Y cannot be influenced by individuals who have the same utilities in both alternatives. The already-dead are the most obvious example of an unaffected subclass, hence the name of the axiom, though any restriction to the dead is not binding, given an anonymity axiom. It is easiest to see what this axiom rules out through an example inspired by Blackorby et al. (2005).

Table 2. Unaffected Utilities Cannot Influence Welfare Judgments of Current Fertility

	Parent	Child 1	Child 2	Euclid
X	50	40		u_e
Y	50	5	50	u_e

Imagine a parent who is going to have a child with a predictably low-quality (but net-pleasurable) life unless significant resources are spent on the child. The mother's options are: X, have only this child and devote significant resources to it; or Y, have a second child and devote standard child-rearing resources to both. For obvious reasons, whether X or Y is chosen, Euclid's utility is unaffected and remains fixed at some unknowable level u_e. The independence of the utility of the dead states that Euclid's utility cannot affect whether X is preferred to Y or vice-versa. This axiom is arguably more defensible than mere addition on the grounds that it doesn't rely on statements of the goodness or badness of individual existence. For this reason, some have found it more plausible.

In a way, this independence axiom is just a further weakening of mere addition. Rather than accepting that the addition of any net-pleasurable life cannot reduce social welfare, independence of the utilities of the dead demands only that such a judgment be independent of the utilities of unaffected individuals. Even with this weakening, however, a surprisingly small class of SWFs satisfies this new set of axioms. To see why, it is instructive to see that AU violates this axiom: If $u_e = 50 \Rightarrow V^{AU}(X) > V^{AU}(Y)$; if $u_e = -10 \Rightarrow V^{AU}(Y) > V^{AU}(X)$. If Euclid had a fantastic life, average utility is already high, and hence new lives pull the average down; if Euclid had a terrible life, adding the same life instead pulls the average up. Therefore, rankings between choices in AU depend on the level of unaffected well-beings. For similar reasons, all welfare functions where the goodness of adding a life to the population is dependent on whether the current quantity or quality of lives will violate this (see "Dropping Separability").

The class of welfare functions satisfying this independence axiom lives within the critical-level generalized utilitarian (CLGU) family. These functions are a generalized version of TU, where the welfare of all individuals above some critical-level α is summed:

$$V^{CLGU} = \sum_{i=0}^{N} (u_i - \alpha)$$

Proponents of CLGU prefer this generalized function to TU because, for a sufficiently large α, the repugnant conclusion is arguably not as repugnant. Rather than many just slightly pleasurable lives being preferred to any given world, a high α guarantees the large-quantity outcome must have sufficiently high levels of well-being for it to be preferable. However, as known from incompatibility results, even a weakening of the repugnant conclusion results in a violation of mere addition or some other axiom. To see this, consider that a life with utility ν, where $\underline{u} < \nu < \alpha$, reduces social welfare despite being a net-pleasurable existence for the individual. Only by setting $\alpha = \underline{u}$ does the CLGU function satisfy mere addition—a ν between α and $\underline{u}$ can no longer exist—but this is merely the special case of TU.

The broader point at this stage is that, when coupled with the original four fixed-population axioms that welfare economists find uncontroversial, acceptance of either mere addition or independence of the utilities of the dead implies that adding sufficiently good lives to a

population is always socially valuable. This is a striking finding. The notion runs counter to widely held intuitions and implies that larger populations with lower standards of living may be better than smaller populations of higher-quality lives. Because it is separability that implies a constant quantity–quality trade-off, separability must be rejected to avoid these implications.

Dropping Separability

The additive separability that characterizes the CLGU family has attractive features but, as noted in "Mere Addition, Independence of the Utilities of the Dead, and Separability," challenges some common ethical intuitions. One response is to conclude that the theories in the literature constitute strong arguments to drop unexamined intuitions.

But it is worth asking about the alternatives: They may prove even better. The most natural non-separable alternative to summing well-being is averaging well-being, as AU does. Because AU is non-separable, it holds that whether or not having a baby improves welfare depends on the number and quality of lives in the Stone Age, as in the example of Table 2. Aside from violating mere addition and separability, AU has its own intuitive challenges for small, rather than large, populations. Namely, it prefers very small populations with high-quality lives to larger populations with even slightly lower average lifetime utility. If humanity had the choice between one excellent generation and a long history with just slightly less excellent generations, AU would recommend the single generation. For these reasons, AU has few (if any) proponents in the modern literature (Ord, 2014).

In attempts to reconcile the attractive properties of TU for small populations and AU for large populations, variable value (VV) theories have been proposed (Hurka, 1983; Ng, 1989; Sider, 1991). These functions nest AU and TU by admitting population into the welfare function, but with diminishing marginal returns:

$$V^{VV}(N,\Theta)= g(N)\times\Theta$$

Here g is a concave function. If g' decreases fast enough, VV functions avoid the repugnant conclusion. However, by mimicking AU for large populations, this function will reject mere addition when N is sufficiently large. And because its judgments on whether adding individuals is socially beneficial depends on the number of past people, it does not satisfy separability. If estimates of ancient Greek populations were upwardly revised, it would reduce the value of new individuals because humanity would be further along the g function than previously believed. In short, explicit efforts to avoid the implications of additive separability introduce their own unintuitive implications.

Separability in AU and VV theories is dropped only across variable-population cases; both theories retain separability in subsets of same-number cases. Others drop separability as a means for jointly incorporating population and inequality concerns. Asheim and Zuber (2014),

for example, proposed and characterized a geometrically rank-weighted social welfare function, called rank-discounted generalized utilitarianism (RDGU), that gives the worse-off people more weight in social evaluation than better-off people. Spears and Stefánsson (2021) showed that the difference between RDGU and CLGU is same-number separability: substituting same-number separability for Asheim and Zuber's novel axiom designed to avoid the repugnant conclusion, but retaining the rest of the axioms that they use to characterize RDGU, yields a characterization of CLGU. Violating same-number separability, which RDGU does, means that the number and welfare of unaffected people far in the past or future cannot be ignored even for everyday economic policy evaluations that do not change the identity or size of the population.

In sum, although dropping or weakening separability axioms has been generally seen as the most promising route to a widely agreeable alternative to the family of CLGU functions, such a consensus alternative has not yet arisen. But even AU, RDGU, and other alternatives identify conditions in which it would be strictly better to add lives to the population, even at a small cost to the welfare of people who would otherwise exist.

Timelessness, Discounting, and Anonymity

It may seem odd that a discussion of discounting has been omitted, even while considering intergenerational utility and leveraging temporal intuitions in earlier arguments. The reason is that the anonymity axiom precludes normative discounting (though one may wish to discount in an expected utility sense to include empirical probabilities of extinction). Anonymity states that, for any two individuals, the social ordering must be indifferent to an exchange of utility between these individuals. It is taken as given that an impartial social order must be invariant to *who* lives the good (bad) lives for a fixed distribution of lifetime utilities. Discounting utilities beyond extinction probabilities violates this: If a current individual's utility were exchanged with a higher quality future life—and the social order discounted future utilities—welfare would increase merely by bringing the good life forward in time. The same is true of summing average intra-generational welfare over time, even without discounting. Generations differ in size, and so the welfare function would prefer exchanging a good life from a larger generation, where it barely influences the generational average, with a bad life from a smaller generation.

Privileging individuals based on when they were born is thought to be as arbitrary as privileging individuals based on where they were born, or other demographic features universally judged as inappropriate for inclusion in a SWF (Greaves, 2017). Although this is a long-standing area of contention in the climate economics literature, recent survey evidence suggests that the majority of discounting experts now believe these rates should be grounded in part by objective, normative premises—like an anonymity axiom—rather than merely the preferences of existing individuals (Drupp et al., 2018).

Recent and interesting work in de la Croix and Doepke (2021) came to the population ethics literature from a new angle that manages to preserve some form of anonymity and discounting simultaneously. The proposed framework relies on probing one's moral intuitions

at the level of the soul (soul-based utilitarianism; SBU), where souls are reincarnated in multiple lifetimes. The social welfare function over variable populations represents a trade-off between the number and the respective qualities of these incarnations. Anonymity is retained at the level of the soul, but because souls may discount the future, anonymity at the incarnation level (i.e., people's lived experiences) fails—souls prefer to move their good experiences ahead in temporal ordering. In fact, the way souls intertemporally discount their time in waiting characterizes the way SBU values population sizes. The authors conclude that only some forms of popular welfare functions can arise as a form of SBU (e.g., AU, TU, and VV can arise, but CLGU cannot).

Summarizing the Theoretical Landscape

Parfit (1984) catalyzed modern research on population ethics by calling on academics to search for "Theory X," a theory that satisfied axioms he found uncontroversial (approximately axioms (1)–(4), mere addition, and non-anti-egalitarianism) and avoided the repugnant conclusion. Incompatibility theorems have since proven that this search was bound to fail. As the landscape is now understood, one must either (a) accept that the repugnant conclusion may be implied by the correct SWF, (b) accept a welfare function that violates mere addition or non-inequality-preferring social aggregation, or (c) give up the existence of an impartial SWF. This final option is unsatisfactory (Parfit, 2011): It would leave people unable to agree that an arbitrarily small sacrifice by one to create arbitrarily large benefits for all other living creatures is beneficial. It would leave one unable to say that it is better to have a climate policy that prevents suffering than one that causes mass suffering, if the policy also inevitably changes the identities of people who exist in 2300.

Between options (a) and (b), the theoretical literature has been dominated by authors seeking the most attractive way of building a welfare function that rejects mere addition in an effort to avoid the repugnant conclusion. Research by Spears and Budolfson (2021) instead demonstrated that all candidate SWFs satisfying minimal aggregative axioms imply some form of the repugnant conclusion. If so, welfarists do not face a choice between "the" repugnant conclusion and some set of axioms, but between different axioms that all lead to some instances of repugnant conclusions. Avoidance of the repugnant conclusion then fails to be interesting or useful if all SWFs imply some form of it. Indeed, for this reason and several others, a consensus statement urges the next generation of population ethics research to de-emphasize avoiding the repugnant conclusion: Zuber et al. (2021) agreed that "avoiding the Repugnant Conclusion is not a necessary condition for a minimally adequate candidate axiology, social ordering, or approach to population ethics," perhaps signaling a reprioritization among population ethicists.

Using Population Ethics in Public Choice: Convergence and Parameterization

Relative to the theoretical literature, the set of papers applying these insights to economic problems is less developed. This section illustrates and summarizes the main findings within this literature, along with guidance on how to apply the insights surveyed above.

Convergence

Despite disagreements in philosophically constructed decision scenarios—such as the cases meant to demonstrate mere addition and the repugnant conclusion—in practice, candidate approaches to population ethics theories agree more than might be anticipated (Budolfson & Spears, 2021). There are empirical reasons for this which this article demonstrates using a stylized example that captures important features of the world and the socioeconomic decisions people face. Then, similar arguments or results the literature has produced in other, more realistic, settings are highlighted. Because AU and TU roughly represent the two most polarized methods for evaluating variable population outcomes (while respecting anonymity) for the moment attention here is restricted to demonstrating when and why these two theories will agree. "Generalized Variable-Population SWFs" extends this discussion of practical applications to the wider class of SWFs.

Suppose one found oneself at the start of an infinite horizon economy with empirical facts not dissimilar to how the world may soon look: average well-being grows at a rate of g_y per year; populations shrink at a rate of d_p per year.[2] Normalizing well-being and populations at 1 in the initial period, any given year has average and total well-being $(1+g_y)^t$ and $(1-d_p)^t \times (1+g_y)^t$, respectively. This normalization is substantive in the case of well-being. These are cardinal values where 0 is the implied value of a neutral life, $\underline{u}$; it is assumed that this is a population of net-positive lives in all periods.

A question that might be asked in such a society is: "should population decline be reduced?" In developed countries with fertility rates already below replacement levels, pro-family policies are being debated and implemented with exactly this goal in mind. (Of course, any actual population policy has important costs, benefits, and interaction with reproductive freedom and rights that are abstracted away from for this toy model.) If these theories—which rely on entirely distinct views regarding the value of new people—agree on questions of population size, it suggests they could agree on a wide range of policy questions.

In this simplified setting, there are closed form solutions to the well-being generated under an AU or TU objective function. These allow one to directly evaluate the welfare effects of reducing d_p.

$$V^{TU}(g_y, d_p) = \sum_{t=0}^{\infty} (1-d_p)^t \times (1+g_y)^t = \frac{1}{d_p - g_y + d_p g_y}$$

$$V^{AU}(g_y, d_p) = \frac{\sum_{t=0}^{\infty} (1-d_p)^t \times (1+g_y)^t}{\sum_{t=0}^{\infty} (1-d_p)^t} = \frac{d_p}{d_p - g_y + d_p g_y} = \frac{1}{1 - {}^{g_y}/_{d_p} + g_y}$$

Both AU and TU agree that social welfare is decreasing in d_p. That is, these distinct criteria agree it is welfare-improving, other things equal, to mitigate population decline. This is unsurprising in the case of TU as it directly values population size. It is less obvious why AU, which is supposedly indifferent to population size, also wants to mitigate population decline. The reason is that life is improving. AU therefore prefers outcomes where the relative weight on future generations is larger; AU prefers the world with more population growth.

This convergence result relies on the assumption that well-being growth is invariant to population growth. Indeed, in Malthusian economies where population growth reduces per capita welfare, disagreement arises between TU and AU over optimal populations (e.g., de la Croix & Doepke, 2021). Modern economic growth theory, however, leaves little reason to believe the world is Malthusian, at least locally. There is instead strong evidence of a causal link running positively from population growth to economic growth (Jones & Romer, 2010). Jones (2020) went as far as to demonstrate theoretically that economic growth may end entirely in the case of long-run population decline. Additionally, Karahan et al. (2019) argued that declining business dynamism has been importantly influenced by slowing population growth; Glover and Short (2020) showed that the aging workforce may have increased monopsony power and can explain a majority of the decline in the labor share; and Weil (1999) and many others studied the decreased average consumption that comes with a larger share of retirees. These lines of research suggest that the stylized example here may even be conservative in its assumption that per capita well-being growth is unaffected by population decline. Although work on the explicit interaction of population ethics and modern growth theory has been limited (see Boucekkine & Fabbri, 2013, for an exploration of this type), existing theory gives reason to believe that this convergence result may hold in many realistic settings, at least on the current margin.

Further evidence in support of this convergence conjecture is the already existing work on various optimal policy problems that directly ask about the robustness to population-based welfare criteria. Lawson and Spears (2018) and Greaves (2019) studied optimal populations in a world with finite resource constraints and find that welfare criteria do not importantly change the optimality conditions. Related, but distinct, work in Scovronick et al. (2017) and Arrhenius et al. (2021) argued that the general direction of optimal emissions is not importantly influenced by choices among population ethics frameworks: climate mitigation should be substantially strengthened, relative to business as usual, whether TU or AU is chosen. Ord (2020) argued that investments to reduce extinction risks are robustly supported across welfare criteria (although Méjean et al. (2020) found a slightly smaller degree of

convergence in an existential risk setting). Finally, Kuruc and McFadden (2021) made a convergence argument running in the opposite direction for animal agriculture: If factory farmed animals have net-negative lives, marginal reductions in their populations increase both total and average (interspecies) well-being.

In sum, even on socioeconomic questions most likely to elicit theoretical disagreement between various population frameworks (existential risk, fertility policy, animal farming, etc.), the literature has found the recommendations of these welfare criteria to be in surprising alignment once real-world facts and constraints are taken into consideration. The intensity of their recommendations for different outcomes may vary, and accordingly, too, the "optimal" outcome. But for the marginal policy changes characterizing public discourse—should CO_2 emissions be reduced, should fertility be incentivized, and so on—these theories will largely agree under the empirical facts of the world. What these convergent recommendations disagree with is the familiar practice of ignoring population size in welfare calculations.

Generalized Variable-Population SWFs

Although convergence results have been identified in the literature, they may not hold in all (or even most) settings. Researchers and practitioners can and should study the sensitivity of their particular policy context to various SWFs. This is especially so in cases where population or distributional concerns exist, the latter having so far been set aside. With the proliferation of SWFs incorporating these respective concerns, and the nonobvious interaction between them, this may seem an intractable task.

A class of SWFs that parsimoniously accounts for a broad set of population and inequality concerns is proposed here (see Arrhenius et al., 2021):

$$w = g(N)\left[h\left(N^{-1}\sum_i f(u_i)\right) - h(f(\alpha))\right]$$

Here again N is population size; u_i are individual utilities; and α is the critical level of individual utility necessary for a life to be socially valuable (where a net-neutral life is normalized at 0, such that for $\alpha > 0$ mere addition is violated). The functions f, g, and h are all non-decreasing. If f and h are the identity function, one has some form of utilitarianism; when $g(N) = N$, this becomes CLGU; if g is also the identity function, it is a form of AU. If f is concave and h is the identity function, one has a form of additively separable prioritarianism (Adler, 2008). Prioritarianism inherits the population considerations of CLGU but is concerned with the distribution of well-being such that there are diminishing returns to individual well-being levels from a social perspective. Finally, if f is concave and $h = f^{-1}$, versions of non-separable egalitarianism are recovered.

This generalized characterization allows researchers to move between these functions with a simple parameterization. As a basic example, if the ranking of two hypothetical policy options appears to depend on the choice of critical level, α can be continuously adjusted to highlight which values support which policies. Likewise, a parsimonious version of prioritarianism has a single parameter that controls the degree of concavity on f, which in turn can be transparently modified. In terms of population concerns, $g(N) = N^{\theta}$ is a weakly increasing function where $\theta = 0$ corresponds to an AU view; $\theta = 1$ corresponds to a TU view; and values within this range correspond to variable value theories. Researchers can design their social welfare function to nest many important welfare criteria, using a small number of parameters that control which considerations dominate. The parameter space can then be mapped into combinations that prefer different policies such that readers can understand the robustness of particular recommendations to a range of ethical views.

The Future of Population Ethics in Economics

Despite the intuitive challenges of ranking outcomes where populations vary, the subliterature of welfare economics concerned with these questions has identified a small set of coherent theories satisfying axioms common to well-studied, fixed-population welfarism. Applications of the resulting welfare criteria highlight that inclusion of population effects importantly influences economic policy valuations, though they happen to do so in much the same direction across a wide range of questions. In light of the ease in which these welfare functions can be incorporated into policy analysis, the authors of this article urge future research to continue exploring their applications within policy evaluation frameworks. Population ethics is ready to move out of the economic textbooks and into practical economic evaluations. Population ethics is ready to be used.

Further Reading

Broome, J. (1985). The economic value of life. *Economica*, *52*(207), 281–294.

Broome, J. (2012). *Climate matters: Ethics in a warming world*. W. W. Norton.

Dasgupta, P. S. (1969). On the concept of optimum population. *Review of Economic Studies*, *36*(3), 295–318.

Dasgupta, P. S. (2019). *Time and the generations: Population ethics for a diminishing planet*. Columbia University Press.

Espinosa, R., & Treich, N. (2021). Animal welfare: Antispeciesism, veganism and a “life worth living.” *Social Choice and Welfare*, *56*(3), 531–548.

Franz, N., & Spears, D. (2020). Mere Addition is equivalent to avoiding the Sadistic Conclusion in all plausible variable-population social orderings. *Economics Letters*, *196*, 109547.

Greaves, H. (2017). Population axiology. *Philosophy Compass*, *12*(11), e12442.

References

Adler, M. (2008). Future generations: A prioritarian view. *George Washington Law Review*, *77*, 1478–1520.

Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, *16*(2), 247–266.

Arrhenius, G., Budolfson, M. B., & Spears, D. (2021). Does climate change policy depend importantly on population ethics? Deflationary responses to the challenges of population ethics for public policy. In M. B. Budolfson, T. McPherson, & D. Plunkett (Eds.), *Philosophy and climate change* (pp. 111–138). Oxford University Press.

Asheim, G., & Zuber, S. (2014). Escaping the repugnant conclusion: Rank-discounted utilitarianism with variable population. *Theoretical Economics*, *9*(3), 629–650.

Blackorby, C., Bossert, W., & Donaldson, D. (1995). Intertemporal population ethics: Critical-level utilitarian principles <https://doi.org/10.2307/2171771>. *Econometrica*, *63*(6), 1303–1320.

Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social choice theory, welfare economics, and ethics*. Cambridge University Press.

Blackorby, C., & Donaldson, D. (1984). Social criteria for evaluating population change. *Journal of Public Economics*, *25*(1–2), 13–33.

Boucekkine, R., & Fabbri, G. (2013). Assessing Parfit's repugnant conclusion within a canonical endogenous growth set-up. *Journal of Population Economics*, *26*(2), 751–767.

Broome, J. (2004). *Weighing lives*. Clarendon Press.

Broome, J. (2018). Efficiency and future generations. *Economics & Philosophy*, *34*(2), 221–241.

Budolfson, M. B., & Spears, D. (2021). Does the repugnant conclusion have important implications for axiology or public policy? In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finnerton-Burns (Eds.), *Oxford handbook of population ethics*. Oxford University Press.

de la Croix, D., & Doepke, M. (2021). A soul's view of the optimal population problem. *Mathematical Social Sciences*, *112*, 98–108.

Deaton, A. (2013). *The great escape: Health, wealth, and the origins of inequality*. Princeton University Press.

Drupp, M. A., Freeman, M. C., Groom, B., & Nesje, F. (2018). Discounting disentangled. *American Economic Journal: Economic Policy*, *10*(4), 109–134.

Glover, A., & Short, J. (2020). *Demographic origins of the decline in labor's share* (BIS Working Papers 874). Bank for International Settlements.

Golosov, M., Jones, L., & Tertilt, M. (2007). Efficiency with endogenous population growth. *Econometrica*, *75*(4), 1039–1071.

Greaves, H. (2017). Discounting for public policy: A survey. *Economics & Philosophy*, *33*(3), 391–439.

Greaves, H. (2019). Climate change and optimum population. *The Monist*, *102*(1), 42–65.

Hurka, T. (1983). Value and population size. *Ethics*, *93*(3), 496–507.

Jones, C. I. (2020). *The end of economic growth? Unintended consequences of a declining population* <https://doi.org/10.3386/w2665> (NBER Working Paper No. 26651). National Bureau of Economic Research.

Jones, C. I., & Romer, P. (2010). The new Kaldor facts: Ideas, institutions, population, and human capital. *American Economic Journal: Macroeconomics*, *2*(1), 224–245.

Karahan, F., Pugsley, P., & Sahin, A. (2019). *Demographic origins of the startup deficit* <https://doi.org/10.3386/w25874> (NBER Working Paper No. 25874). National Bureau of Economic Research.

Kuruc, K., & McFadden, J. (2021). *Monetizing the externalities of animal agriculture: Insights from an inclusive welfare function* <https://sites.utexas.edu/pwi/%20files/2021/11/AgInclusiveWelfare.pdf> (Population Wellbeing Initiative White Paper No. 5). The University of Texas at Austin.

Lawson, N., & Spears, D. (2018). Optimal population and exhaustible resource constraints. *Journal of Population Economics*, *31*(1), 295–335.

Méjean, A., Pottier, A., Fleurbaey, M., & Zuber, S. (2020). Catastrophic climate change, population ethics and intergenerational equity. *Climatic Change*, *163*(2), 873–890.

Narveson, J. (1973). Moral problems of population. *The Monist*, *57*(1), 62–86.

Ng, Y. K. (1989). What should we do about future generations? Impossibility of Parfit's Theory X. *Economics & Philosophy*, *5*(2), 235–253.

Ord, T. (2014). Overpopulation or underpopulation? In I. Goldin (Ed.), *Is the planet full?* (pp. 46–60). Oxford University Press.

Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.

Parfit, D. (1984). *Reasons and persons*. Oxford University Press.

Parfit, D. (2011). *On what matters*. Oxford University Press.

Roberts, M. (2020). Population ethics, the mere addition paradox, and the structure of consequentialism <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/

9780190905323.001.0001/oxfordhb-9780190905323-e-20>. In D. Portmore (Ed.), *The Oxford Handbook of Consequentialism*. Oxford University Press.

Scovronick, N., Budolfson, M. B., Dennig, F., Fleurbaey, M., Siebert, A., Socolow, R. H., Spears, D., & Wagner, F. (2017). Impact of population growth and population ethics on climate change mitigation policy. *Proceedings of the National Academy of Sciences*, *114*(46), 12338–12343.

Sider, T. (1991). Might theory X be a theory of diminishing marginal value? *Analysis*, *51*(4), 265–271.

Spears, D., & Budolfson, M. B. (2021). Repugnant conclusions. *Social Choice and Welfare*, *57*(3), 567–588.

Spears, D., & Stefánsson, O. H. (2021). Additively-separable and rank-discounted variable-population social welfare functions. *Economic Letters*, *203*, 109870.

Weil, D. (1999). Population growth, dependency, and consumption. *American Economic Review: Papers and Proceedings*, *89*(2), 251–255.

Zuber, S., Venkatesh, V., Tännsjö, T., Tarsney, C., Stefánsson, O. H., Steele, K., Spears, D., Sebo, J., Pivato, M., Ord, T., Ng, Y. K., Masny, M., MacAskill, W., Lawson, N., Kuruc, K., Hutchinson, M., Gustafsson, J., Greaves, H., , . . .Forsberg, L. Asheim, G. (2021, April 13). What should we agree on about the repugnant conclusion? *<https://doi.org/10.1017/S095382082100011X>* *Utilitas*, 1–5.

Notes

1. Prioritarianism, for example, is total utilitarianism adjusted with concavity over lifetime utilities. Prioritarianism fits this description because it retains the TU additive quality with respect to all individuals with a net-pleasurable life, but this SWF prefers increasing the utility of individuals further down the well-being distribution, whereas TU is indifferent to the distribution of utility.

2. This setup leads to a zero population (after rounding) in finite time, hence the finite welfare values. Although perhaps unrealistic, it is useful for illustrative purposes and is isomorphic to the case of stable long-run populations with a positive probability of extinction.

Related Articles

Public Finance and Soft Budgets

Poverty and Social Policy in the United States

Financial Protection Against Medical Expense

Returns to Health Spending in Low- and Middle-Income Countries

Consumer Debt and Default: A Macro Perspective